第17講 データの分析

§1 データの整理と代表値

要点整理 と 公式

要点1 データとその整理

「変量」… 人や物の特性を数値的に表すもの　（ex）身長，気温，テストの点数

「データ」… 変量の測定値の集まり

（ex）次のデータ1は，あるクラスの生徒10人のハンドボール投げの結果である。

データ1　（単位m）

13，21，15，18，18，16，17，15，19．18

データを整理する方法として，「度数分布表」や「ヒストグラム（柱状グラフ）」などがある。

要点2 データの代表値

「代表値」… データ全体の特徴を1つの数値で表したもの。主な代表値として

「　　　」「　　　」「　　　」などがある。

「平均値」… 変量 x についての n 個の値が x_1，x_2，……，x_n のとき，

これらの和を個数 n で割ったもの。　　で表す。

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \cdots\cdots + x_n)$$

（ex）データ1の平均値　　は

$$\overline{x} = \frac{1}{10}(13 + 21 + \cdots\cdots + 18) = \frac{170}{10} = 17.0\ (\mathrm{m})$$

「最頻値」… データにおいて，最も個数が多い値。モードともいう。度数分布表に

整理したときは，度数の最も多い階級の「階級値」（階級の中央の値）。

（ex）データ1の最頻値は18 m

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

「中央値」… データを大きさの順(小さい順，大きい順)に並べたとき，中央にくる値。

メジアンともいう。データの個数が偶数のときは，中央の2つの値の平均。

(ex) データ1を小さい順に並べると

13，15，15，16，17，18，18，18，19，21

よって中央値は，　　番目と　　番目の値の平均値で $\frac{17+18}{2}=17.5$ (m)

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

1

［1］　次のデータの平均値，中央値を求めよ。

(1)　32，40，31，45，80

(2)　45，41，39，50，44，45

［2］　右の表は，あるクラスの生徒 27 人のソフトボール投げの結果である。

(1)　中央値が含まれる階級を求めよ。

(2)　最頻値を求めよ。

階級（m）	度数（人）
12以上14未満	2
14 ～ 16	4
16 ～ 18	9
18 ～ 20	5
20 ～ 22	7
計	27

Point Pickup

代表値（平均値・最頻値・中央値）

平均値 … $\overline{x} = \dfrac{1}{n}(x_1 + x_2 + \cdots\cdots + x_n)$

最頻値 … 最も個数の多い値，または度数が最も大きい階級の階級値

中央値 … 大きさの順に並べたとき，中央の位置にくる値

データの個数が奇数個　→　中央の値

データの個数が偶数個　→　中央の 2 つの値の平均

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

｛解答｝

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

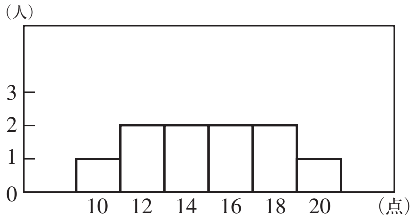
§2 ▶ データの散らばり

要点整理 と 公式

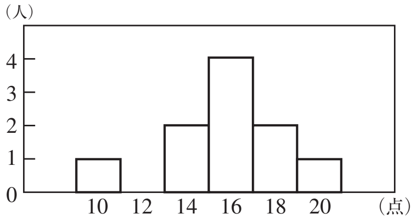
要点3 データの範囲

次の 2 つの図は，あるクラスの 10 人に 2 つのゲーム A，B を行い，その得点をヒストグラムに表したものである。

ゲームA



ゲームB



「範囲」… データの最大値から最小値を引いた差。

これが大きいとデータの散らばりの度合いが大きいと考えられる。

上のゲームA，ゲームB の範囲はともに，$20-10=10$ (点)である。

よって，範囲だけでは「散らばりの違いはない」となる。

最大値や最小値に極端に離れた値があると，範囲はそれに大きく影響を受ける。

要点4 四分位数と四分位範囲・四分位偏差

「四分位数」… データを大きさの順に並べたとき，「　　　　」する位置にくる

「　　つの仕切り」となる数。この 3 つの四分位数は，小さい方から

「第　　四分位数」「第　　四分位数」「第　　四分位数」といい，

順に　　，　　，　　と表す。

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

（ex）奇数個のデータ 1，2，3，4，5，6，7，8，9 について

（1，2，3，4，）　5，　（6，7，8，9）

↑
中央値 5
これを　　とする。

1，2，3，4　　　　　　6，7，8，9

↑　　　　　　　　　　↑

中央値 $\dfrac{2+3}{2}=2.5$　　　　中央値 $\dfrac{7+8}{2}=7.5$

これを　　とする。　　　　これを　　とする。

「四分位範囲」… 第 3 四分位数から第 1 四分位数を引いた数 ＿＿＿＿＿

データの中央およそ 50% の範囲。

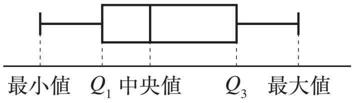
「四分位偏差」… 四分位範囲を 2 で割った数

＿＿＿＿＿

データが中央値(Q_2)の周りに集中しているほど，四分位範囲や四分位偏差は小さい傾向にあり，逆に四分位範囲や四分位偏差が大きいと，散らばりの度合いが大きいと考えられる。

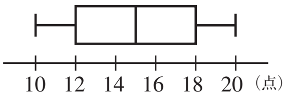
要点❺ 箱ひげ図

「箱ひげ図」… 最小値，Q_1，中央値，Q_3，最大値を，箱と線(ひげ)を用いて1つの図に表したもの。

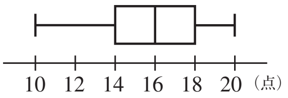


前ページのゲームA，ゲーム B の得点の箱ひげ図は，それぞれ次のようになる。

ゲーム A



ゲーム B



このように，箱ひげ図は，データの分布のおおまかな様子を表している。よって，複数のデータの分布を比較することに適している。

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

2

次のデータは，ある本について書店 A，B における 1 日の販売数 10 日分を，小さい順に並べたものである。

書店 A … 0，0，3，3，4，5，7，8，8，9

書店 B … 2，3，5，5，6，7，7，8，8，9

(1) 各データの四分位数，四分位範囲，四分位偏差を求めよ。

(2) 2 つのデータの箱ひげ図を並べてかけ。

(3) 2 つのデータの散らばりの度合いを比較せよ。

Point Pickup

散らばりの度合い

範囲の大きさや四分位範囲の大きさから，複数のデータの散らばりの度合いを比較することができる。

範囲 …（最大値）−（最小値）

四分位範囲 … $Q_3 - Q_1$

四分位偏差 … $\dfrac{Q_3 - Q_1}{2}$

3 つの四分位数 Q_1，Q_2，Q_3 は，データを 4 等分する位置に来る値である。

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

{ 解答 }

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

§3 分散と標準偏差

要点整理 と 公式

要点6 偏差と分散，標準偏差

ここでは，データの値すべてに関してその散らばり度合いをみる方法の１つとして，平均値との差を考える。

データの個々の値と平均値との差が大きいほど，その値は平均から離れている。

変量 x についての n 個の値を x_1，x_2，……，x_n とし，その平均を $\overline{x}$ とする。

各値の平均値からの差 $x_1-\overline{x}$，$x_2-\overline{x}$，……，$x_n-\overline{x}$ を，それぞれの平均値からの「　　　」という。各値が平均値より大きければ偏差は　　，平均値より小さければ偏差は　　，平均値と同じ値であれば偏差は　　となる。

この偏差の平均値は

$$\frac{1}{n}\{(x_1-\overline{x})+(x_2-\overline{x})+\cdots\cdots+(x_n-\overline{x})\}=\frac{1}{n}\{(x_1+x_2+\cdots\cdots+x_n)-\overline{x}\times n\}$$

$$=\frac{1}{n}(x_1+x_2+\cdots\cdots+x_n)-\overline{x}=\overline{x}-\overline{x}=0$$

つまり常に０となり，偏差の平均値を用いて散らばりの度合いを見ることはできない。

ここで，偏差を「　　　」した値を考えると，その値は０以上となり，また各値が平均値より離れるほどその値は大きくなる。よってこの平均値の大小は，データの散らばり度合いの大小を表す１つの指標となる。この偏差の２乗の平均値を「　　　」といい，「　　」で表す。

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

また，分散の正の平方根 $\sqrt{s^2}=$ 　　を「　　　　　」という。

■分散と標準偏差

分散　$s^2=\dfrac{1}{n}\left\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\right\}$

標準偏差　$s=\sqrt{s^2}=\sqrt{\dfrac{1}{n}\left\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\right\}}$

（ex）ある小テストの得点のデータが 5，6，7，8，9 のとき

平均値は $\overline{x}=\dfrac{1}{5}(5+6+7+8+9)=7$（点）

x	5	6	7	8	9
$x-\overline{x}$	−2	−1	0	1	2
$(x-\overline{x})^2$	4	1	0	1	4

分散は　$s^2=\dfrac{1}{5}\left\{(-2)^2+(-1)^2+0^2+1^2+2^2\right\}=\dfrac{1}{5}(4+1+0+1+4)=2$

標準偏差は $s=\sqrt{s^2}=\sqrt{2}\fallingdotseq 1.4$（点）

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

3

次のデータは，8 人の生徒が 2ヶ月の間に読んだ本の冊数 x（冊）である。

1，3，6，5，8，1，6，2

(1)　このデータの平均値 $\overline{x}$ を求めよ。

(2)　このデータの分散 s^2 を求めよ。

(3)　このデータの標準偏差 s を求めよ。

Point Pickup

分散と標準偏差

変量 x の n 個の値が x_1，x_2，……，x_n，その平均値が $\overline{x}$ のとき

分散　$s^2=\dfrac{1}{n}\{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots\cdots+(x_n-\overline{x})^2\}$　…（偏差）^{2} の平均値

標準偏差　$s=\sqrt{s^2}$　…$\sqrt{(分散)}$

平均値が複雑な値のときの分散

平均値が複雑な値のときには，分散を求める際の偏差やその 2 乗の計算に時間を要する。よって，次の式を用いるとよい。

分散　$s^2=\overline{x^2}-(\overline{x})^2$　…（x^2 の平均値）−（x の平均値）^{2}

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

{ 解答 }

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

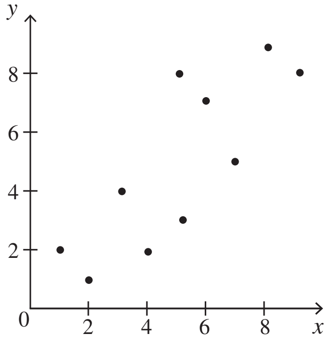
§4 ▶ データの相関

要点整理 と 公式

要点 7 散布図と相関関係

ある 2 つの変量の間に関係があるかを知りたいとき，その 2 つのデータの値を座標として 1 つの図に表す方法がある。

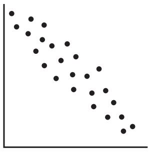
x	8	4	3	5	2	1	6	9	5	7
y	9	2	4	8	1	2	7	8	3	5



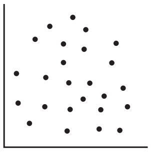
右の図は，x，y の値の組を座標$(x,\ y)$として座標平面上にとったものである。このように，2 つの変量からなるデータを座標平面上に図示したものを「　　　　　」という。

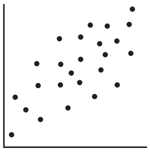
2 つの変量からなるデータにおいて，一方が増えると他方も増える傾向にあるとき，2 つの変量には「　　の相関がある」という。また，一方が増えると他方が減る傾向にあるとき，2 つの変量には「　　の相関がある」という。正の相関も負の相関もみられないとき，「相関がない」という。

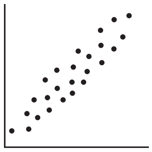
散布図において，点の分布が直線に近づくほど「相関が　　　」といい，広く散らばっているほど「相関が　　　」という。


負の相関（強い）

負の相関（弱い）


相関がない


正の相関（弱い）


正の相関（強い）

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

要点8 相関係数

2つの変量 x，y からなるデータの組 $(x_1,\ y_1)$，$(x_2,\ y_2)$，……，$(x_n,\ y_n)$ について，

偏差の積 $(x_k - \overline{x})(y_k - \overline{y})$ の平均値を「共分散」といい，「 s_{xy} 」と表す。s_{xy} が正のときは正の相関があり，s_{xy} が負のときは負の相関がある。s_{xy} が 0 に近いときは，相関はない。

また，共分散 s_{xy} を，x，y の標準偏差 s_x，s_y の積 s_xs_y で割ったものを，x，y の「相関係数」といい，「 r 」で表す。

$-1 \leqq r \leqq 1$ であり，r の値が 1 に近いほど正の相関が強く，-1 に近いほど負の相関関係が強い。また，0 に近いほど，相関関係は弱くなる。

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

4

次の表は，2つの変量x，yからなるデータである。

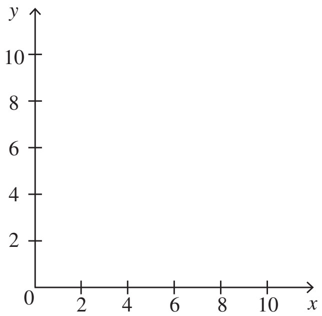
x	3	2	7	6	8	4	9	1	3	7
y	8	9	5	4	1	6	2	7	5	3

(1)　散布図をかき，x，yの間に相関があるか調べよ。また，相関がある場合は，正，負のどちらの相関関係があると考えられるかを答えよ。

(2)　(1)の散布図から，このデータのxとyの相関係数rに最も近いものを次の①～④から1つ選べ。

①　-0.31　　②　-0.88　　③　0.25　　④　0.92

＜散布図＞



Point Pickup

散布図と相関関係

散布図における点の分布が

右上がりの直線に近い　→　正の相関がある

右下がりの直線に近い　→　負の相関がある

どちらの傾向もない　→　相関がない

相関係数 r

相関係数rの値が

1に近い　→　強い正の相関がある

-1に近い　→　強い負の相関がある

0に近い　→　相関が弱い（ない）

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

{ 解答 }

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

第17講　確認テスト

1　次のデータは，10 人の生徒の小テストの得点である。

18，16，12，6，18，12，16，14，18，20　　（単位は点）

平均値は［アイ］点，中央値は［ウエ］点，最頻値は［オカ］点である。

2　次のデータ について，下の問いに答えよ。

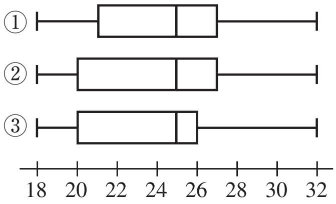
18，19，21，23，25，26，26，28，32

(1)　第 1 四分位数は Q_1 = ［キク］，第 2 四分位数は Q_2 = ［ケコ］，

第 3 四分位数は Q_3 = ［サシ］である。

(2)　このデータを箱ひげ図に表したものは右の図の［ス］である。

［ス］にあてはまるものを①～③から選べ。



3　変量 x のデータ 11，13，7，11，13，11 について，平均値は $\bar{x}$ = ［セソ］，分散は s^2 = ［タ］，

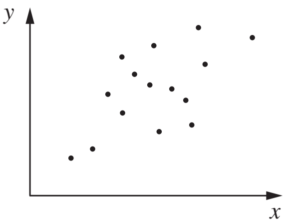
標準偏差は　s = ［チ］である。

4　右の散布図から，2 つの変量 x，y には［ツ］といえる。

［ツ］にあてはまるものを，下の①～③から選べ。

①　正の相関関係がある　　　②　負の相関関係がある

③　相関はみられない



©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

{ Note }

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

第17講　データの分析

1

［1］（1）平均値 $\frac{228}{5}=45.6$　　中央値 40

（2）平均値 $\frac{264}{6}=44$　　中央値 $\frac{44+45}{2}=44.5$

［2］（1）16m 以上18m 未満

（2）$\frac{16+18}{2}=17$（m）

2

（1）書店 A のデータについて

第 1 四分位数は $Q_1=3$（冊）

第 2 四分位数は $Q_2=\frac{4+5}{2}=4.5$（冊）

第 3 四分位数は $Q_3=8$（冊）

四分位範囲は $Q_3-Q_1=5$（冊），四分位偏差は $\frac{Q_3-Q_1}{2}=\frac{5}{2}=2.5$（冊）

書店 B のデータについて

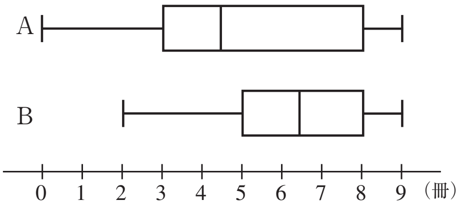
第 1 四分位数は $Q_1=5$（冊）

第 2 四分位数は $Q_2=\frac{6+7}{2}=6.5$（冊）

第 3 四分位数は $Q_3=8$（冊）

四分位範囲は $Q_3-Q_1=3$（冊），　四分位偏差は $\frac{Q_3-Q_1}{2}=\frac{3}{2}=1.5$（冊）

（2）



（3）書店 A のデータの範囲は　9 − 0 = 9（冊），書店 B のデータの範囲は　9 − 2 = 7（冊）

よって，範囲，四分位範囲ともに書店 A のデータの方が大きいので，

書店 A のデータの方が散らばりの度合いが大きいと考えられる。

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

3

(1)　$\bar{x} = \dfrac{1}{8}(1+3+6+5+8+1+6+2) = \dfrac{32}{8} = 4$(冊)

(2)　$s^2 = \dfrac{1}{8}\{(-3)^2+(-1)^2+2^2+1^2+4^2+(-3)^2+2^2+(-2)^2\} = \dfrac{48}{8} = 6$

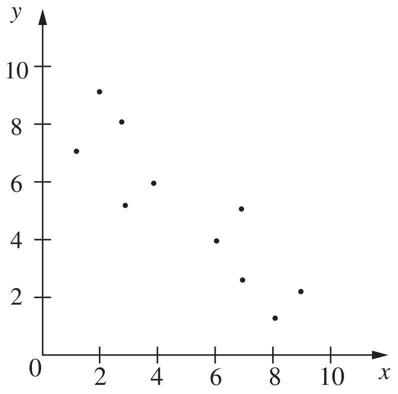
(別解)　$\overline{x^2} = \dfrac{1}{8}(1^2+3^2+6^2+5^2+8^2+1^2+6^2+2^2) = \dfrac{176}{8} = 22$

よって，$s^2 = \overline{x^2} - (\bar{x})^2 = 22 - 4^2 = 6$

(3)　$s = \sqrt{s^2} = \sqrt{6} = 2.449\cdots \fallingdotseq 2.5$(冊)

4

(1)



負の相関関係がある

(2)　強い負の相関関係が見られる。よって，相関係数 r の値は -1 に近い。

したがって，②

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

第17講　確認テスト

1　次のデータは，10 人の生徒の小テストの得点である。

18，16，12，6，18，12，16，14，18，20　　（単位は点）

平均値は［アイ］点，中央値は［ウエ］点，最頻値は［オカ］点である。

2　次のデータ について，下の問いに答えよ。

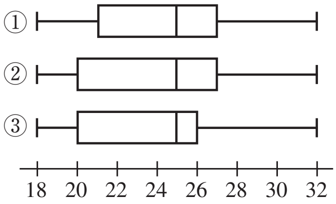
18，19，21，23，25，26，26，28，32

(1)　第 1 四分位数は Q_1 = ［キク］，第 2 四分位数は Q_2 = ［ケコ］，

第 3 四分位数は Q_3 = ［サシ］である。

(2)　このデータを箱ひげ図に表したものは右の図の［ス］である。

［ス］にあてはまるものを①～③から選べ。



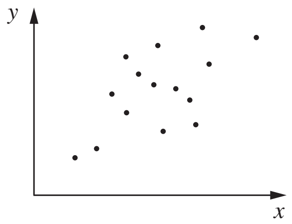
3　変量 x のデータ 11，13，7，11，13，11 について，平均値は $\bar{x}$ = ［セソ］，分散は s^2 = ［タ］，標準偏差は　s = ［チ］である。

4　右の散布図から，2 つの変量 x，y には［ツ］といえる。

［ツ］にあてはまるものを，下の①～③から選べ。

①　正の相関関係がある　　　②　負の相関関係がある

③　相関はみられない



©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。

｛解答｝

1　ア　1　　イ　5　　ウ　1　　エ　6　　オ　1　　カ　8

2　(1)　キ　2　　ク　0　　ケ　2　　コ　5　　サ　2　　シ　7

(2)　ス　②

3　セ　1　　ソ　1　　タ　4　　チ　2

4　ツ　①

©RECRUIT HOLDINGS
本サービスに関する知的財産権その他一切の権利は著作権者に帰属します。
また本サービスに掲載の全部または一部につき無断複製・転載を禁止します。